基于 CNN-SVM 和转发树的微博事件情感分析

- 1. 中国科学院声学研究所 语音声学与内容理解重点实验室 北京 100190;
- 2. 中国科学院新疆理化技术研究所 新疆民族语音语言信息处理实验室 乌鲁木齐 830011;
- 3. 中国科学院大学 北京 100190

涂曼姝 1 张艳 1 颜永红 1,2,3

摘要 基于微博的情感分析近几年获得了广泛的关注,但是通常是对微博上的电影或者产品的评论。 我们的研究目标是针对微博上的热点事件的分析,将这些事件的评论分为正向或者负向,将微博用户 的评论进行分类有利于辨别公众对于这个事件的普遍看法。本文提出了一个基于卷积神经网络的混合 模型: CNN-SVM,用于对事件的评论进行分类。其次,根据微博用户的转发行为,我们提出了一种 新的数据结构——转发树,用以解决评论中的一些分类混乱情况。实验结果表明 CNN-SVM 确实提 高了情感分类的正确率,新提出的转发树结构在对真实世界的情感倾向性的逼近中也是十分有效的。

关键词:情感分析,CNN-SVM,转发树

中图分类号: G35

Sentiment Classification Based on CNN-SVM and Repost Tree to Weibo Events

- 1. The Key Laboratory of Speech Acoustics and Content Understanding Institute of Acoustics, Chines e Academy of Sciences, Beijing 100190, China;
- 2. Xinjiang Laboratory of Minority Speech and Language Information Processing, Xinjiang Technical Institute of Physics and Chemistry, Urumchi 830011, China;
- 3. Chinese Academy of Sciences University of Chinese Academy of Sciences, Beijing 100190, China

TU Manshu¹ ZHANG Yan¹ YAN Yonghong^{1,2,3}

基金项目: 本研究受国家自然科学基金(11461141004, 61271426, U1536117, 11504406, 11590770-4), 国家863计划 (2015AA016306), 中科院与博后基金会联合基金(2015LH0041)的资助。

作者简介:涂曼姝(1991-),博士研究生,研究方向:倾向性分析,Email:tumanshu@hccl.ioa.ac.cn;张艳(1973-),博士,副研究员,研究方向:自然语言理解、数据挖掘、社会计算、大数据分析;颜永红(1967-),博士,研究员,研究方向:语音信号处理、语音识别、大数据分析。

探索与研究

DISCOVERY AND RESEARCH

doi:10.3772/j.issn.2095-915x.2017.03.010

Abstract Sentiment classification on weibo recently attracted research community the widespread attention. Most previous works focused on weibo comments about movies or products. In contrast, our study aims at the hot events on weibo. Comments of the events are considered either positive or negative on behalf of the user's attitude to these events. Classification of user attitude helps to identify the opinions of the general public. In this paper, we put forward an innovative method based on convolution neural network, termed as CNN – SVM, to classify the comment. In addition, according to the forwarding behavior of users, we put forward a new data structure, repost tree for dealing with ambiguity in the comments. Extensive experiments demonstrated that the CNN–SVM method effectively improved the accuracy of events sentiment classification. The new data structure showed to be effective on steering the classification results towards real world sentiment tendency.

Keywords: Sentiment classification, CNN-SVM, repost tree

1 引言

随着网络的发展,微博逐渐改变了人们了解新闻的方式,微博用户可以即时的讨论不同的新闻事件。根据文献 [1],截止 2015 年 9 月,微博月活跃用户数 (MAU) 已经达到 2.12 亿人,比 2014 年同期增长了 48%,微博已经变成了公众表达自己态度的重要平台之一。因此通过对微博的舆论监督来检测哪些事件可能会演变成暴力恐怖事件变得十分重要。在本文中,我们主要研究分析公众对事件的情感倾向性(积极或是消极),我们希望通过结合以前的情感分析方法用来提高分类的正确率,与此同时我们利用微博的数据特征通过一些逻辑运算来获得公众对于热点事件的真实情感倾向性。

基于微博的情感分析方法总体上来说分为 机器学习和基于词典的方法。基于词典的方法 简单来说就是使用情感词典对每一个文本进行 打分从而对文本进行分类,由于情感词典的领 域局限性,基于词典的方法效果往往没有机器 学习的方法好。 Pang 使用电影评论数据来比较不同的机器学习方法(朴素贝叶斯、最大熵分类和支持向量机 SVM)在情感分类上的效果,实验表明SVM 在情感分类上的结果最好^[2]。SVM 最早是由 Cortes 在 1995 年提出的^[3]。它通常被用来解决小样本,非线性和高维特征的模式识别问题^[4]。当测试的数据集的特征比较稀疏时 SVM也拥有较好的分类能力^[6]。SVM 也被广泛的应用于人脸识别,说话人识别和图像处理上^[5]。Zainuddin等人通过大量的实验证明了特征选择在情感分析中扮演很重要的角色^[7]。他们通过选择不同的特征提取方法并使用 SVM 作为分类器进行实验。

传统的机器学习方法在情感分类这一任务上往往需要大量的人工设计特征。Collobert 和Weston提出了一个基于深度神经网络的方法,它可以在先验知识有限的情况下自主学习相关特征。这种方法可以实现多任务学习并且在没有人工设计特征的前提下提高 NLP 任务的表现^[8]。Liang 使用递归神经网络(RNN)来发现相关特征^[9],他们的工作也可以自主发现情感分析

SENTIMENT CLASSIFICATION BASED ON CNN-SVM AND REPOST TREE TO WEIBO EVENTS

相关的特征。卷积神经网络(CNN)通过特征 和滤波器的卷积得到特征的更高维度的表示。 它最初用在计算机视觉方面的[10], 之后 CNN 模型被证明在 NLP 领域也有不错的效果[11]。 Kim 训练了一个在词向量上有一层卷积层的简 单 CNN 模型 [12], 他将谷歌新闻上的 1000 亿个 词通过一个非监督的语言模型训练得到词向量, 并用这种词向量与其他论文中常用的词向量分 别作为 CNN 的输入来比较不同词向量的使用对 分类效果的影响,结果表明这种提前训练的词 向量在许多NLP的分类任务上都有较好的表现。 与传统的文本任务不同, 由于微博文本长度限 制在140个字符,对于微博的倾向性分析是极 具挑战性的。Santos 等人^[13]提出使用两层的卷 积神经网络分别抽取句子和词语级别的信息, 这个方法在电影评论和 twitter 信息的分析上有 较好的表现。Tang 等人通过联合情感词嵌入特 征(SSEW)和一组手工标注特征作为输入提出 了一种深度学习网络来解决情感分类问题[14]。 Johnson 受到图像拥有 RGB 和 CMYK 通道的启 发,将句子视为图像将单词视为像素[15]来解决 情感分类问题。

综上所述,基于 CNN 和 SVM 的方法在情感分析领域都是目前流行的算法。将二者结合的方式虽然在图片的识别上取得了较好的效果 [16],然而,据我们所知,将其结合的方法还没有在自然语言处理领域(NLP)被使用过。同时通过调研我们发现,马莹莹在其硕士毕业论文中所提到的对于微博转发情况的考虑一方面是对于转发行为的预测,另一方面是对其转发内容的倾向性的预测 [17],对于已经发生的转发行为可能会产生的倾向性反转的情况并没有

纳入考虑。为了解决这些问题,我们提出了将CNN与SVM结合构建一个新的分类器模型,我们同时也提出了一种新的数据结构——转发树,它可以在进行倾向性分析时将上下文考虑其中。转发树可以对树的每一个节点的倾向性进行逻辑运算从而判断该语句的真实倾向性。我们的实验表明了在转发树结构下的CNN-SVM得到的倾向性结果更接近真实世界的倾向性结果。

本文包括下面主要内容:第二章将描述 CNN-SVM 的模型和转发树模型。第三章描述 了数据的情况和数据预处理。我们也将在这一 章节展示实验结果和分析结果。在第四章我们 会对已有工作进行总结并对未来的研究方向进 行展望。

2 模型

2.1 CNN-SVM

我们提出的模型如图 1 所示。这个模型是在 Kim 的 CNN 模型的基础上进行的改进 [11]。在模型中,CNN 用来进行特征提取,SVM 作为分类器用来进行倾向性分类。这个模型的工作方式如下。

模型的输入是矩阵 $S \in R^{(s \times n)}$ 。符号 n 表示每一个句子中每一个词都用 n 个训练好的向量表示: $[v_1, \cdots, v_i, \cdots, v_n]$ 。符号 s 表示每一个句子包含 s 个单词: $[w_1, \cdots, w_i, \cdots, w_s]$ 。模型的卷积层有许多不同种类的滤波器 $F \in R^{(m \times n)}$,m 表示滤波器的宽度,宽度不同就表示不同的滤波器。在卷积时,取与滤波器宽度相同的句子片段作为

DISCOVERY AND RESEARCH

doi:10.3772/j.issn.2095-915x.2017.03.010

输入,则一个句子的第 i 个向量矩阵通过滤波器得到结果的计算公式如公式 1 所示:

 $c_i = (S * F)_i = \sum_{ki} (S_{[i-m+1:i,:]} \otimes F)_{ki}$ (公式 1) k 表示的是第 k 层卷积层, \otimes 是卷积运算符, $S_{[i-m+1:i,:]}$ 是宽度为 m 的矩阵块。每一个句子的特征矩阵 S 通过一个滤波器经过卷积运算都

会得到一个特征映射 $c \in R^{(1\times(S-M+1))}$ 。为了形成 更丰富的特征,每一个滤波器都有 p 个卷积核。 假设我们有 t 种滤波器,那么一个句子就会得 到 p×t 个特征映射 $C:[c_{f_1}, \cdots, c_{f_{f_p}}, c_{f_1}, c_{f_p}]$ 。Maxpooling 是一个采样层,每一组特征映射 c 输出 最大值: $c_{pool}: R^{1\times(s-m+1)} \to R^{1\times 1}$ 。

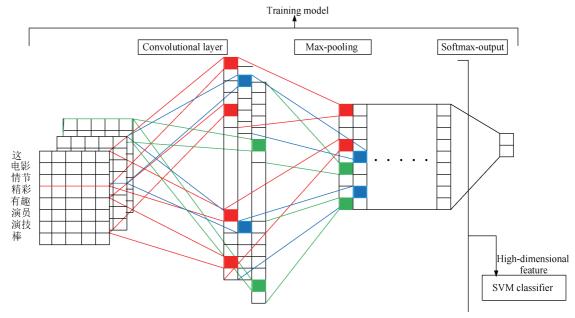


图1 在这个图中有三种不同的滤波器。红色表示的滤波器的宽度是3,蓝色表示的滤波器的宽度是4,绿色表示的是2.不同的矩形表示的是不同的句子。每一个句子都要经过这三种滤波器。每一个行向量表示一个单词。在卷积层的最后,每一列都表示一个句子通过一种滤波器后的结果。若每一个滤波器都含有p个卷积核,则每一种特征映射都会产生p个单元特征。这些单元特征经过采样生成p个高维的特征值。

现在我们得到了高维的特征表示: $D:[c_{pool_{11}},\cdots,c_{pool_{1p}},c_{pool_{1p}}]$ 。采样层的输出与一个 soft-max 层全连接。它用公式 2 来计算句子的标签:

$$\begin{split} p(y = j \mid c_{pool}, B) &= soft \max_{j} (c_{pool}W + B) \\ &= \frac{e^{c_{pool}w_{j} + b_{j}}}{\sum_{k=1}^{K} e^{c_{pool}w_{j} + b_{k}}} \quad (\text{ $\angle \vec{x}$ } 2) \end{split}$$

 w_k 和 b_k 是第 k 层的权重向量和偏移向量。Softmax 层通过比较预测的标签值和真实的标签值来调整 CNN 模型。当训练集的正确率在 CNN

上表现稳定时,我们将经过采样层得到的测试 集的高维的特征表示提取出来,放到 SVM 中。 然后我们训练 SVM 模型直到它正确率最高。保 持我们训练得到的参数,然后将测试集的数据 送入到模型中用来分类。

2.2 转发树

微博用户在浏览微博的过程中会对自己感 兴趣的内容转发并评论。这种评论行为往往是 对其转发的内容进行的评论,因此若其转发的

SENTIMENT CLASSIFICATION BASED ON CNN-SVM AND REPOST TREE TO WEIBO EVENTS

内容本身是带有情感倾向性的话,则评论的倾向性是在内容的倾向性基础上的倾向性。例如,当某事件发生后,用户对该事件表达了负向情感。然后另一个用户转发这个负向情感,并对这个转发的信息表达了正向情感。("哈哈,同意//出轨一年不自省,刚被揭露就道歉。对于这种行为我只能说四个字:不可原谅!")。这意味着第二个用户对这个事件是负向的情感。如果我们仅仅分析语句"哈哈,同意",分类器将会把这个句子分成正向。基于这种原因我们提出了转发树这种结构。该结构将转发的情况分成三种。

- 1. 只转发不评论:与其转发的微博拥有相同的倾向性。
- 2. 转发并带有负向情感: 其倾向性与其转发的微博对评论的事件的倾向性相反。
- 3. 转发并带有正向情感: 其倾向性与其转 发的微博对评论的事件的倾向性相同。

由于微博在每一次转发过程中都会保留路径,这为构建转发树提供了便利。这种路径包含昵称和微博内容:"@+昵称:+微博内容+//"("@新浪四川:→_→//@成都微吧:→_→//@李伯伯:我们让你跌倒?")。我们是使用转发树这种结构来调整通过分类器的得到的倾向性分类的结果。转发树结构包括昵称,微博内容和时间戳。

转发树的结构如图 2 所示。我们将转发树按照第一层节点的子节点数的多少降序排序。 在圆中的数字是根据分类算法得到的倾向性分类结果。然后转发树就可以进行它的逻辑运算步骤。

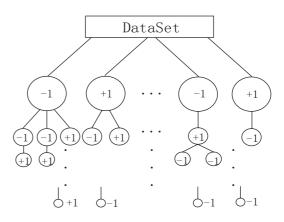


图2 不同尺寸的圆圈代表不同的层级。一个微博越早发布,那么它的圆圈越大。在圆圈里面或者外面的数字表示的是经过分 类所发得到的倾向性分类结果。

逻辑运算遵循以下步骤。我们定义所有在一个分支上的节点状态为: $Branch_{node} = n_1, \dots, n_j, \dots, n_n$, 节点 n_i 的值只能是+1或者-1。 $Node_n$ 的逻辑运算的结果被视为该节点真正的情感倾向性。在对树进行逻辑运算时采用自下而上的方法。它的计算公式如公式3所示:

$$Node_n = \prod_{i=1}^N n_i$$
 (公式 3)

3 实验

3.1 数据集和预处理

我们使用了三个测试集。他们都是从新浪微博中抽取得来。

表1 数据集

Dataset Name	Training Data	Test Data	
MN	1600		
WZ	23120	5780	
PARIS	480	120	

蒙牛数据集: 这个数据集是由 COAE-2013

DISCOVERY AND RESEARCH

doi:10.3772/j.issn.2095-915x.2017.03.010

提供的已标注数据。该数据由对蒙牛事件的微博评论组成。该事件是蒙牛集团被查出在牛奶中添加三聚氰胺导致一些孩子成了大头娃娃的事件。我们是用 MN 作为蒙牛事件的缩写。

文章数据集:这个数据集是由一个中国著名演员文章出轨后相关的微博评论组成,它包括一百万未标注数据。我们使用WZ作为文章数据集的缩写。

巴黎数据集:这个数据集是关于 2015 年 11 月 13 日发生在法国巴黎的暴力恐怖袭击事件的数据集。这个数据集包括 3 万未标注的评论。我们使用 Paris 作为巴黎暴恐事件的缩写。

我们使用 3.2 所描述的方法对 MN 和 Paris 数据集进行标注。三个数据集都使用以下步骤进行预处理:删除网址链接,使用 ICTCLAS 分词,去掉停用词,去掉词数少于 2 的句子。数据集具体的情况如表 1 所示。训练集和测试集都拥有相同数量的正向和负向句子。所有数据集中训练集的数量都是测试集的 4 倍。

3.2 远程监督和特征提取

3.2.1 情感符

越来越多的微博用户习惯使用表情符来表达自己的观点。微博自带许多表情符,这些表情符中的一些有非常明显的情感倾向性,因此我们使用这些表情符作为标签,具体做法如 Go et al. 论文所示 [18]。在微博中每一个表情符都有一个中文词语对应,并且有许多表情都对应这一个中文词。在我们获得的评论文本中,表情符都是以"[哈哈]"这种形式出现的。我们将所有带有明显情感倾向性的表情转换成 +1 或

者-1,+1表示的是正向情感,-1表示的是负向情感,同时生成一个表情符词典。表2列出了一些词典中的例子。如果一个句子包含了多于一种类型(既有正向又有负向)的表情符,我们就删除它,因为多样的表情符意味着这个句子可能同时表达了正向和负向情感。

3.2.2 远程监督

当一件事情发生并且变成热门事件后,对于这件事件的评论会特别的多。对这些评论进行一一的标注时间代价太大,因此我们使用远程监督的方法标注数据 [18]。我们选择所有带有词典中的表情符的评论,讲这些评论打上 +1 或者 -1 的标签。然后我们将训练集中的句子删除表情符。去掉表情符的目的是希望模型更加关注句子的文本特征(词带特征,Word2vec,TF-IDF)。如果我们不处理表情符,就会对分类算法产生一个负面的影响。但是在测试集中我们保留表情符以便保持句子的原始信息。

3.2.3 特征提取

我们使用公开可获得的 Word2vec 特征向量,这个向量是用 3000 亿微博语句训练得到的。它包括 1043631 个词,其中每个词都由 200 维特征进行表示。该向量特征已经被用在多个

SENTIMENT CLASSIFICATION BASED ON CNN-SVM AND REPOST TREE TO WEIBO EVENTS

NLP 任务上并且展示了最佳的性能 [19]。没有被包含的词将会被随机分配一些向量。

3.3 其他方法

我们使用 SVM 作为基线。除此之外我们也是用了 Kolchyna et al. 的方法构建情感词典 [20]。构建得到的情感词典有 2850 个负向词和 2907个正向词,这些词的值在 -1 和 +1 之间。我们将每一个句子表示为: [w1,…,w1,…w1],句子的得分公式如公式 4 和公式 5 所示。如果词典中不包含某个词,那么这个词的分值为 0。

$$S_{score} = \left(\sum_{i=1}^{S} w_i\right) \setminus num_{w_i \neq 0} \tag{公式 4}$$

$$S_{score} = \begin{cases} positive, s_{score} > 0\\ negative, s_{score} < 0 \end{cases}$$
 (公式 5)

我们也使用了深度信念网络(DBN)和受限玻尔兹曼机(RBM)用来比较性能^[21,22]。同时为了使实验有说服力,我们也使用了简单的CNN模型。所有的机器学习的方法都使用了五折交叉验证方法。

3.4 实验和结果分析

表 3 是使用不同的方法对三个数据集进行 分类的结果。它显示了 CNN-SVM 的正确率 和其他方法的正确率。我们首先比较三个不 同的数据集在不同的分类器下的结果,我们 发现所有的方法都在 MN 数据集上有最好的 表现。原因是 MN 数据集是标准的数据集, 在这个数据集中所有的句子都是完整和连贯 的。

在这三个数据集上, DBN 和 RBM 的效果 都不理想的。一方面 DBN 是由 RBMs 组成的, 因此当 RBM 的效果不好的时候 DBN 的效果也 不会好。另一方面, RBM 并不是很适合对不标 准的文本进行分类。在三个数据集上 CNN 都 有较好的表现。与 SVM, RBM 和 DBN 这三个 分类器中最好的结果相比,它的分类准确率在 第一个数据集中提高了7.86%, 在第二个数据 集中提高了15.75%,在第三个数据集中提高了 4.69%。我们提出的 CNN-SVM 与 CNN 相比表 现出了更好的性能。相较于 CNN, CNN-SVM 的分类准确率在第一个数据集中提高了1.32%, 在第二个数据集中提高了5.58%,在第三个数 据集中提高了2.65%。第一个数据集的正确率 提升不如第二个和第三个数据集, 可能是因为 在第一个数据集上 CNN 已经表现了很好的性 能,提升空间本来就不大了。在第三个数据集 中 SVM 的性能比在另外两个数据集中要好,因 为 SVM 更适合小的数据集。

表3 分类正确率结果

数据	情感词典	SVM	RBM	DBN	CNN	CNN-SVM
MN	75.18%	84.28%	84.17%	84.25%	91.96%	93.28%
WZ	61.35%	55.73%	56.23%	54.38%	71.98%	77.56%
Paris	62.29%	70.16%	56.86%	54.55%	74.85%	77.50%

我们发现情感词典的方法在第一个数据集中的表现最差,但是在第二个和第三个数据集

中都有一个较好的表现。原因是我们使用情感符标记了这两个数据集,这些表情符转化的中

DISCOVERY AND RESEARCH

doi:10.3772/j.issn.2095-915x.2017.03.010

文词可能在我们的情感词典中并且拥有较高的权重值。

对于转发树,我们只使用 Paris 的数据做实验,这个数据集拥有 14470 条转发微博。在网站上,(这个网站报道了人民日报关于巴黎暴恐这个事件公众的情感倾向性的调查),我们将这个调查结果作为公众对于

该事件的真实倾向性的反应。在图 3 中,经过 CN-SVM 分类后的倾向性结果与真实的倾向性相差 20%,但是经过转发树的逻辑运算,倾向性的值就更加接近真实情况了。根据图 3 的结果,CNN-SVM 与转发树的结合拥有最好的效果,与真实的倾向性仅相差 7.46%。

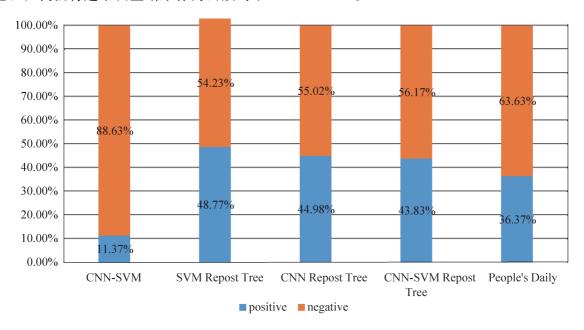


图 3 SVM与转发树相连接和CNN与转发树相连接、CNN-SVM与转发树相连接和真实的情感倾向性作比较。

4 总结

在本文中我们提出了 CNN-SVM 混合方法,并且与其他方法进行了比较,实验结果表明这种方法在微博事件的情感倾向性分析中取得了最好的效果,它可以大幅度提高事件的情感倾向性的分类正确率。除此之外,我们提出的新的数据结构 -- 转发树,可以进一步调整由 CNN-SVM 得到的倾向性结果,使结果更接近真实世界的倾向性,它在某种程度上加强了CNN-SVM 的分类结果。转发树也可能含有误

差,因为一些人转发微博时的评论有可能是针对的原始事件而不是其转发的微博,我们会在未来的工作中考虑这种情况。

参考文献

[1] 新浪微博数据中心. 《2015 年度微博用户发展报告》 (2015) [EB/OL]. [2015-12-30]. http://www.useit.com.cn/thread-10921-1-1.html

[2] Pang B, Lee L, Vaithyanathan S. Thumbs up?: Sentiment Classification using Machine Learning Techniques[C]// ACL 2002 Conference on Empirical Methods in Natural Language Processing. Association

- for Computational Linguistics, 2002: 79-86.
- [3] Cortes C, Vapnik V. Support-vector Networks[C]// Machine Learning. 1995: 273-297.
- [4] 刘霞, 卢苇. SVM 在文本分类中的应用研究 [J]. 计算机教育, 2007(2): 72-74.
- [5] 祁亨年. 支持向量机及其应用研究综述 [J]. 计算机工程, 2004, 30(10): 6-9.
- [6] Joachims T. Text categorization with Support Vector Machines: Learning with Many Relevant Features[M]. Machine Learning: ECML-98. Springer Berlin Heidelberg, 1999: 137-142.
- [7] Joachims T. Text Categorization with Support Vector Machines: Learning with Many Relevant Features[M]. Machine Learning: ECML-98. Springer Berlin Heidelberg, 1999: 137-142.
- 8] Collobert R, Weston J. A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning[C]// International Conference. DBLP, 2008: 160-167.
- [9] 梁军, 柴玉梅, 原慧斌, 等. 基于深度学习的微博情感分析 [J]. 中文信息学报, 2014, 28(5): 155-161.
- [10] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based Learning Applied to Document Recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [11] Kim Y. Convolutional Neural Networks for Sentence Classification[J]. Eprint Arxiv, 2014.
- [12] Yih W T, Toutanova K, Platt J C, et al. Learning Discriminative Projections for Text Similarity Measures[J]. 2011: 247-256.
- [13] Santos C N D, Gattit M. Deep Convolutional Neural Networks for Sentiment Analysis of Short

- Texts[C]// International Conference on Computational Linguistics. 2014.
- [14] Tang D, Wei F, Qin B. Coooolll: A Deep Learning System for Twitter Sentiment Classification. In Proceedings of the 8th International Workshop on Semantic Evaluation. 2014: 208-212.
- [15] Johnson R, Zhang T. Effective Use of Word Order for Text Categorization with Convolutional Neural Networks[J]. Eprint Arxiv, 2014.
- [16] Niu X X, Suen C Y. A novel hybrid CNN-SVM Classifier for Recognizing Handwritten Digits[J]. Pattern Recognition, 2012, 45(4): 1318-1325.
- [17] 马莹莹.微博用户转发行为及情感预测研究 [D]. 哈尔滨:哈尔滨工业大学, 2015.
- [18] Go A, Bhayani R, Huang L. Twitter Sentiment Classification using Distant Supervision[J]. Cs224n Project Report, 2009.
- [19] Iyyer M, Enns P, Boyd-Graber J, et al. Political Ideology Detection Using Recursive Neural Networks[C]// Meeting of the Association for Computational Linguistics. 2014: 1113-1122.
- [20] Kolchyna O, Souza T T P, Treleaven P, et al. Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination[M]. Handbook of Sentiment Analysis in Finance. arXiv, 2015.
- [21] Hinton G E, Osindero S, Teh Y W. A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [22] Zhou S, Chen Q, Wang X. Fuzzy Deep Belief Networks for Semi-supervised Sentiment Classification[J]. Neurocomputing, 2014, 131(9): 312-322.